# RESEARCH PAPERS

*Acta Cryst.* (1996). D**52**, 611–622

## An Eye Lens Protein–Water Structure: 1.2 Å Resolution Structure of γB-Crystallin at 150 K

V. SAGAR KUMARASWAMY,[a] PETER F. LINDLEY,[b] CHRISTINE SLINGSBY[c] AND IAN D. GLOVER[a]

[a]*Physics Department, Keele University, Keele, Staffordshire ST5 5BG, England,* [b]*CCLRC Daresbury Laboratory, Daresbury, Warrington, Cheshire WA4 4AD, England, and* [c]*Crystallography Department, Birkbeck College, Malet Street, London WC1E 7HX, England*

(*Received 11 May 1995; accepted 16 October 1995*)

### Abstract

γB-crystallin is a structural protein of the eye lens with a role in the maintenance of an even distribution of protein and water over distances around the wavelength of light, preserving lens transparency. The structure of the 174-residue bovine protein has already been determined at room temperature to 1.47 Å resolution. By flash freezing the protein crystals, data have now been collected to a nominal resolution limit of 1.2 Å as radiation damage was essentially eliminated. The protein–water model has been refined against this data using the program *RESTRAIN* converging to an *R* factor of 18.5% with all data. Atomic positions are clearly indicated in the electron-density maps. Discrete bimodal disorder has been visualized for a few side chains. Out of a total of 498 water molecules present in the crystal asymmetric unit, 394 have been modelled and refined at unit occupancy. The solvent structure is extremely well ordered with an average *B* value of 23.4 Å$^2$. Partially occupied sites have been identified where disorder in the protein induces concomitant disorder in the local solvent structure. The solvent structure covers 97% of the solvent-exposed surface of the protein in the crystal. 126 water molecules are distributed in second and higher hydration shells. There are networks of hydrogen-bonded solvent extending up to 64 molecules in a network, comprising trimers and tetramers as well as five- and six-membered water-ring structures. The hydration of the protein surface is dominated by arginine and aspartate side chains. Extensive cages of highly ordered solvent molecules are also observed around exposed non-polar groups.

### 1. Introduction

The major components of eye lenses are α-, β-, γ- and δ-crystallins. α-Crystallins are large polymeric assemblies that also function as small heat-shock proteins (Bloemendal & de Jong, 1991; Klemenz, Fröhli, Steiger, Schäfer & Aoyama, 1991; Horwitz, 1992) whereas β- and γ-crystallins are structurally related and form a superfamily of oligomers and monomers (Wistow & Piatigorsky, 1988; Lubsen, Aarts, & Schoenmakers, 1988). β- and γ-crystallin polypeptide chains are folded into two similar β-sheet domains, each domain being composed of two Greek-key motifs (Blundell *et al.*, 1981; Bax *et al.*, 1990). The closely related family of monomeric γ-crystallins comprise around 170 amino acids in which the two domains are linked covalently by a connecting peptide. γ-Crystallins are predominantly found in the relatively dehydrated core regions of lenses where the refractive index is high and this leads to a degree of hardness (Slingsby, 1985). The molar ratio of protein to water in the mature rat eye lens core (Philipson, 1969) is similar to that in crystals of γB-crystallin where it is 862 mg ml$^{-1}$. In the avian lens, where rapid changes of shape are needed to alter the accommodation range, a soft hydrated lens is required and the γ-crystallins have been replaced by the large, oligomeric, α-helical δ-crystallin, the X-ray structure of which has recently been determined (Simpson *et al.*, 1994). There is almost no protein turnover in the core regions of lenses (Wannermacher & Spector, 1969) and therefore the γ-crystallins must have great stability in terms of their tertiary structures and their interactions with other cell components including water. Young mammalian lenses are liquid-like and exhibit cold cataract whereby the γ-crystallins separate into phases of unequal density on cooling (Benedek *et al.*, 1979, Broide, Berland, Pande, Ogun & Benedek, 1991; Tardieu, Vérétout, Krop & Slingsby, 1992). This core region of the human lens frequently turns opaque in senile cataract when γ-crystallins become cross-linked with other lens components (Harding, 1991).

High-resolution X-ray crystallography allows the visualization of the preferred hydration sites on proteins (Karplus & Faerman, 1994) and the results have recently been reviewed (Frey, 1993). γB-crystallin has 174 amino-acid residues and a molecular weight of 20 967 Da. The structure has previously been refined at 1.47 Å resolution using data collected with synchrotron radiation at ambient temperature (Najmudin *et al.*, 1993). A total of 230 water molecules surrounding the protein were located and refined. These studies highlighted




ISSN 0907-4449 

Table 1. *Statistics for the assessment of data quality, given for the data processed to 1.2 Å resolution and including the data collected to 2 Å resolution*

| $D_{min}$ (Å) | $\langle I \rangle$ | $\sigma, (I)$ | $N_{meas}$ | $N_{indep}$ | $I > 3\sigma(I)$ | % Completeness | $R$ |
|---|---|---|---|---|---|---|---|
| 3.79 | 1657 | 231 | 7313 | 1693 | 96.3 | 98 | 0.058 |
| 2.68 | 1493 | 182 | 16211 | 2897 | 99.2 | 97 | 0.061 |
| 2.19 | 880 | 131 | 23740 | 3662 | 98.1 | 97 | 0.077 |
| 1.90 | 756 | 144 | 24568 | 4213 | 97.7 | 94 | 0.095 |
| 1.70 | 399 | 42 | 15509 | 4539 | 96.4 | 94 | 0.060 |
| 1.55 | 253 | 27 | 16661 | 4900 | 94.1 | 92 | 0.068 |
| 1.43 | 166 | 23 | 17239 | 5144 | 89.4 | 91 | 0.092 |
| 1.34 | 123 | 22 | 17549 | 5388 | 83.9 | 90 | 0.125 |
| 1.26 | 97 | 23 | 17758 | 5540 | 77.8 | 89 | 0.163 |
| 1.20 | 80 | 24 | 16578 | 5317 | 69.4 | 88 | 0.200 |
| Overall | | | 173126 | 43293 | | | 0.077 |

Grand total of output reflections (including those measured only once) 45977

some shortcomings of crystallographic data collection at ambient temperatures. Firstly, it was apparent that the crystals would diffract to considerably higher resolution than 1.47 Å if radiation damage could be prevented. Secondly, these studies had indicated the partial oxidation and formation of a disulfide bond between residues 18 and 22 in the γB-crystallin molecule. Furthermore, whilst a large number of water molecules had been located and refined at 1.47 Å, including 150 hydrogen bonded to the protein, some 498 water molecules per asymmetric unit are expected.

In order to address these problems a data set was collected at a cryogenic temperature, 150 K, where radiation damage is virtually abolished. Indeed two complete data sets at high and low resolution were collected from the same crystal with no sign of time-dependent radiation damage. The low-resolution, $d_{min} = 2.0$ Å, data set has been used in a preliminary refinement which confirms the overall identity of the molecular structure at low and ambient temperatures. It also showed, unambiguously, the absence of a disulfide linkage between Cys18 and Cys22 (Lindley *et al.*, 1993). At cryogenic temperatures thermally enabled atomic vibrations are reduced, enhancing the signal-to-noise ratio of high-resolution data. This approach enables considerable extension in data resolution to at least 1.2 Å and in common with other reports (Teeter, Roe & Heo, 1993) has enabled the visualization of statistical disorder in the crystal. The major benefit in this case has, however, been the location and refinement of much more of the solvent structure. Here, we report the application of cryocrystallography to determine a complete protein–water surface structure, as well as higher hydration layers, of γB-crystallin in crystals where the overall level of hydration is similar to that of lenses.

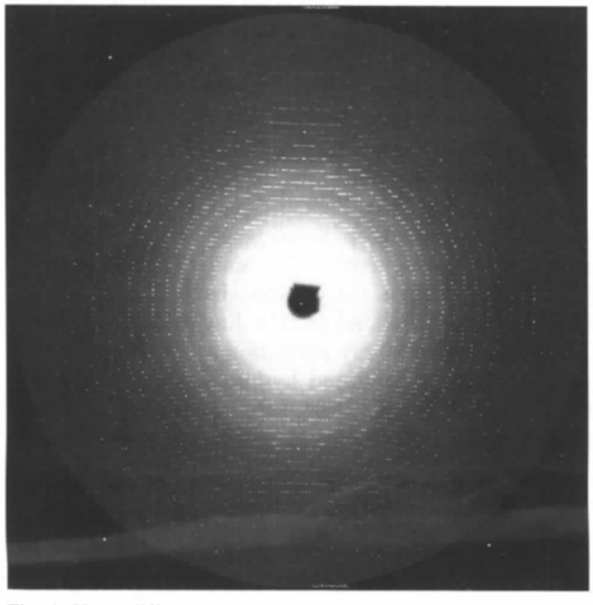


Fig. 1. X-ray diffraction photograph of γB-crystallin.

## 2. Experimental

### 2.1. *Crystals and crystallographic data collection*

Crystals of bovine γB-crystallin were grown at low salt concentration, 0.05 *M* sodium hydrogen phosphate, pH 6.8, in the presence of dithiothreitol, as described previously (Carlisle, Lindley, Moss & Slingsby, 1977) and cooled to 150 K during data collection, using crystals less than four weeks old. The crystals have a low solvent content, 36%(*v*/*v*), similar to that of crambin crystals which have 32%(*v*/*v*) solvent (Teeter, Roe & Heo, 1993). A crystal was transferred to an oil droplet, adhered to a fine glass fibre and introduced to a nitrogen gas stream at 150 K (Hope, 1988). The oil coating acts merely to prevent crystal dehydration during transfer and not as a cryo-protectant. Large crystals (0.5 × 0.5 × 1.0 mm) showed significant loss in crystallinity upon freezing, rendering them useless for data collection. Smaller crystals diffracted well although with an approximate twofold increase in mosaic spread compared to that at room temperature. All diffraction data

Table 2. *Model refinement at 1.2 Å resolution*

| | |
|---|---|
| Resolution range (Å) | 12–1.2 |
| *R* factor (%), all data | 18.4 |
| Correlation coefficient | 95.71 |
| No. of reflections | 45905 |
| No. of restraints | 3816 |
| Total No. of observations | 49721 |
| No. of protein atoms | 1486 |
| No. of water molecules | 394 |
| No. of positional parameters | 5640 |
| No. of individual isotropic *B* values | 1889 |
| Occupancy groups | 4 |
| Overall scale and *U* | 2 |
| Total No. of parameters | 7532 |
| Data-to-parameter ratio, *Q* | 6.61 |

were collected using station 9.6 at the Daresbury SRS using an incident wavelength of 0.89 Å. In a typical diffraction pattern (Fig. 1) the resolution limit of 1.18 Å is set by the data-collection geometry, indicating that higher resolution data collection will be possible.

A 1.2 Å data set was recorded from a single crystal of γB-crystallin, space group $P4_12_12$, dimensions $0.2 \times 0.2 \times 0.8$ mm. Data were recorded using an Arndt–Wonacott camera and packs of three CEA Reflex 25 X-ray film, with blank films interspersed between the first and second pairs of films. The crystal was mounted with the *c* axis parallel to the camera rotation axis and a 45° sweep of data, $\Delta\varphi = 1.0°$, using exposure times of 750 s per degree. This was followed by the collection of a 2.0 Å data set from the same crystal with shorter exposure times and larger $\Delta\varphi$ (Lindley *et al.*, 1993).

The data were processed using the *MOSFLM* suite of programs based on the *MOSCO* suite (Nyburg & Wonacott, 1977). The films were digitized using a Joyce–Loebl Scandig 3 microdensitometer. Due to the proximity of the spots along the *c* axis, the films were scanned using a 25 μm² raster, optical density 0.0–2.0 units. The refined cell dimensions $a = b = 56.43$ (0.05), $c = 97.12$ (0.05) Å represent a shrinkage of 4.8% in unit-cell volume compared to ambient temperature. The effect of the increase in the observed mosaicity of the crystal was adequately explained using a value of 0.3° for the mosaic spread. Profile-fitted intensities within one pack were scaled together using the program *ROTCOR* and corrected for Lorentz and polarization corrections with *PASCAL*.

No corrections for absorption were applied. Scaling and merging of the data were performed using the Fox & Holmes (1966) algorithm as implemented in the

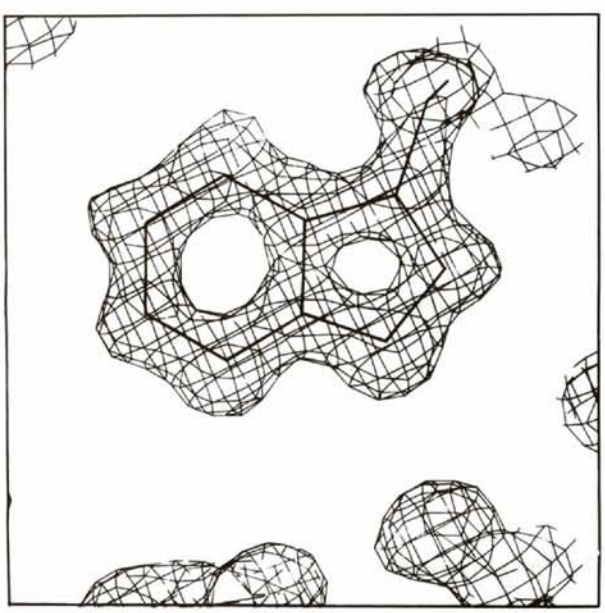
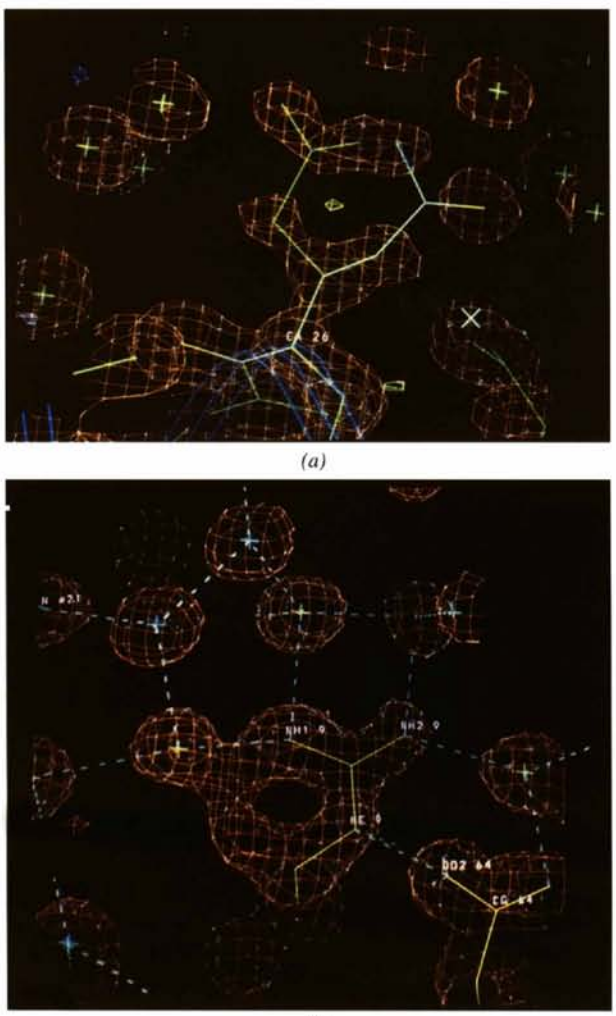


Fig. 2. The quality of the map; electron density at 1σ level of Trp42.



(*a*)

(*b*)

Fig. 3. Bimodal disorder: electron density, at the 1σ level, showing alternate conformers of (*a*) the side chain of Gln26, (*b*) Arg9 where side-chain unoccupied sites are occupied by water molecules; NH1 and NH2 atoms contribute to five-membered hydrogen-bonded rings with water molecules.

*CCP*4 (Collaborative Computational Project, Number 4, 1994) programs *ROTAVATA* and *AGROVATA*. A total of 173 126 measurements were processed to yield 45 977 unique reflections with an overall $R_{merge}$ of 6.7%. Because of the exposure times used, the low-resolution shells of data were completely saturated, resulting in poor completeness at low resolution. The 2.0 Å data set (Lindley *et al.*, 1993) was subsequently merged into this data set. Data processing and statistics are summarized in Table 1. The low-temperature approach has enabled the collection of high-quality high-resolution data from γB-crystallin. At the limit of resolution the data are over 80% complete and 70% of the outer-shell reflections have $I > 3\sigma(I)$.

### 2.2. *Refinement*

Stereochemically restrained reciprocal-space least-squares refinement using the program *RESTRAIN* (Driessen *et al.*, 1989) was initiated using starting phases calculated from the low-temperature 2.0 Å model, excluding all solvent molecules. Three cycles of rigid-body refinement were carried out at 2.0 Å resolution to ensure accurate positioning of the model in the unit cell. The model was refined to convergence using data in the 8.0–1.5 Å shell (24 459 reflections) and data resolution extended in three stages, 8–1.4 Å (25 948 reflections), 8–1.2 Å and finally 12–1.2 Å (45 905 reflections) to include as many low-resolution terms as possible. One of the features of high-resolution data is the overdeterminancy or $Q$ factor of the least-squares refinement. The $Q$ factor for reflections only is over 6 offering the possibility of unrestrained refinement using an isotropic thermal model. However, when restraints were relaxed the model geometry rapidly deteriorated reflecting the fact that although 1.2 Å resolution is high for proteins it is still some way removed from atomic resolution. An *ad hoc* procedure of several cycles of refinement of low restraint weights followed by tighter restraints was adopted throughout the refinement. Even with tight restraints the contribution of structure-factor terms to the overall residual was over 95%.

During the course of the refinement extensive use of electron-density maps was made, examining both $2|F_o| - |F_c|$ and $|F_o| - |F_c|$ maps and model building using *FRODO* (Jones, 1978) on an Evans and Sutherland PS300 graphics system. Water molecules were added to the coordinate list according to the following criteria: (*a*) the water molecule must make at least one stereochemically reasonable hydrogen bond, (*b*) it should have well defined density above $1\sigma$ in a $2|F_o| - |F_c|$ map, (*c*) density above $3\sigma$ in an $|F_o| - |F_c|$ map. The final model included 394 water molecules placed according to these criteria, all refined at unit occupancy.

Examination of electron-density maps and rebuilding of the structure after cycles of refinement gave evidence of discrete disorder in some exposed side chains. These became particularly apparent towards the end of the refinement. The side chains of residues Arg9, Ser19, Gln26 and Asp64 all showed discrete conformational disorder. Early in the refinement the side chains of Lys2, Leu155 and Met160 had poorly defined electron density, suggesting multiple conformers. The electron density for these groups, however, improved during the course of the refinement allowing modelling in new conformations. Discrete disorder was modelled where possible as distinct conformers and occupancies of the alternative sites refined. In most of these cases the alternative (non-filled site) was found to be occupied by water molecules, introducing local disorder into the solvent structure. The C-terminal residues, Phe173 and Tyr174 showed poor side-chain density at the end of refinement, suggesting multiple conformers and, therefore, disorder in this region of the molecule at this temperature.

Table 3. *Solvent hydrogen-bond interactions*

(*a*) Distribution of solvent molecules based on number of protein contacts

| No. of bonds to protein | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| No. of water molecules | 126 | 154 | 76 | 34 | 5 |

(*b*) Analysis of hydrogen-bond distances in water molecules making four hydrogen bonds

| No. of water molecules | 31 | 58 | 97 | 126 |
|---|---|---|---|---|
| No. of hydrogen bonds to solvent | 1 | 2 | 3 | 4 |
| No. of hydrogen bonds to protein | 3 | 2 | 1 | 0 |
| Hydrogen-bond distance (Å) | 2.91 | 2.87 | 2.84 | 2.82 |

(*c*) Analysis of number of $B$ values in second and higher hydration shells as a function of number of observed water contacts

| No. of protein contacts | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| No. of water molecules | 13 | 22 | 34 | 57 |
| Average $B$ value (Å$^2$) | 27.1 | 23.4 | 19.3 | 12.8 |
| Average distance (Å) | 2.83 | 2.84 | 2.81 | 2.82 |

### 2.3. *Solvent-accessibility calculations*

In order to calculate the surface accessibility of the protein a complete shell of atoms within 10 Å of a molecule of γB-crystallin was generated using the crystallographic symmetry operators. The surface-accessible area of all atoms in the protein molecule was then calculated using the rolling-sphere algorithm of Richardson, as implemented in the *CCP*4 program *SURFACE*, using a sphere of 1.4 Å radius. The accessible areas were calculated for the entire molecule, specific residues or hydrophobic/hydrophilic atoms from a summation of the individual atomic accessible areas (C, S atoms hydrophobic, N, O atoms hydrophilic).

## 3. Results and discussion

### 3.1. *The electron-density map and model*

The refinement using *RESTRAIN* converged at an $R$ factor of 18.4% for all data in the 12–1.2 Å shell. The

final model comprises 1474 protein non-H atoms and 394 solvent water molecules. In the model there are four additional side-chain conformers with occupancies less than one. The important refinement statistics are shown in Table 2.

The protein structure is essentially that as reported by Lindley *et al.* (1993) but with the models of disorder as discussed above. Fast cooling of protein crystals normally induces a shrinkage in unit-cell volume, (Hartmann *et al.*, 1982; Teeter, Roe & Heo, 1993) and in this case the reduction in unit-cell volume of 4.8% is accommodated by slight changes in intermolecular packing and surface side-chain conformations. After superposition of the two models resulting from the 1.47 Å resolution, room-temperature refinement and this refinement, using the program *LSQKAB*, (Collaborative Computational Project, Number 4, 1994) r.m.s. deviations in coordinates were calculated. For main-

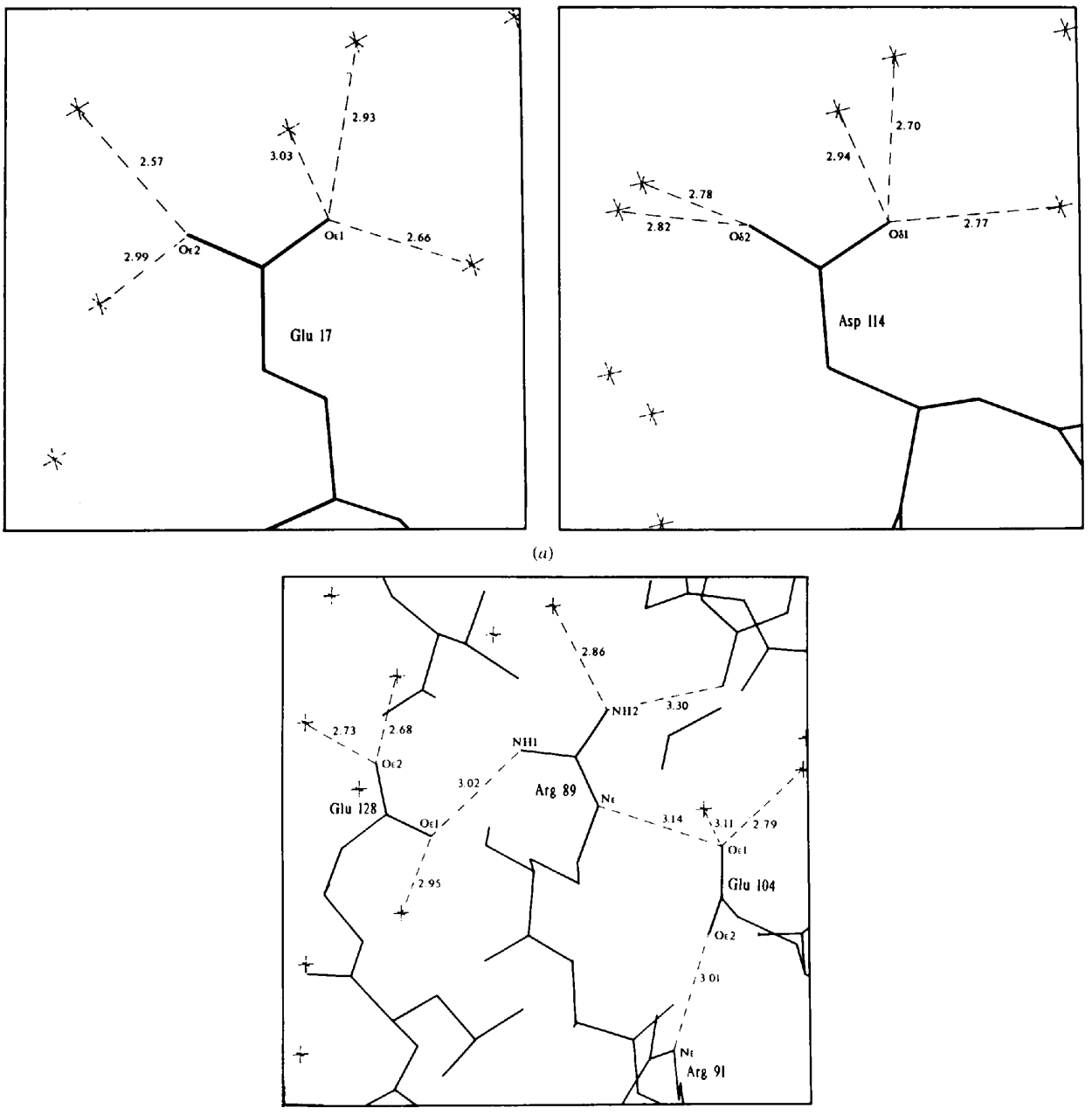
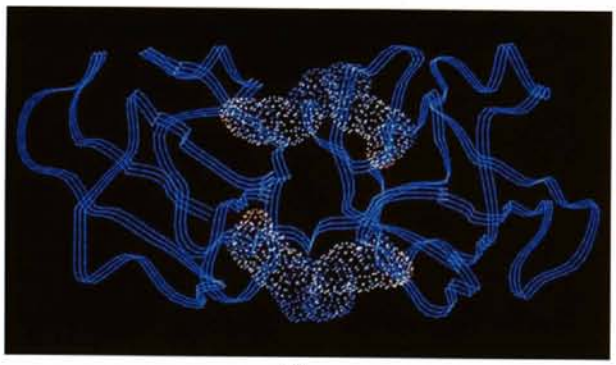



Fig. 4. Isolated charges and ion pairs in γB-crystallin. (*a*) Isolated side chains of residues of Glu17 and Asp114. (*b*) The hydration around the intramolecular ion pair (Glu104–Arg89). Numbers refer to hydrogen-bonding distances.



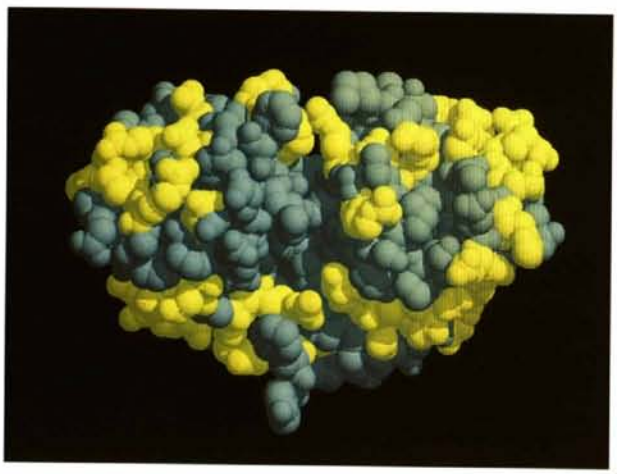
(a)

chain atoms, side-chain atoms, and all protein atoms the r.m.s. deviations were 0.31, 0.99 and 0.75 Å, respectively. The model stereochemistry was analysed using the *PROCHECK* suite (Laskowski, MacArthur, Moss & Thornton, 1993) and all geometric parameters fell within the expected range for a protein structure determined at this resolution. All residues except glycines fell in allowed regions of a Ramachandran plot and expected patterns of side-chain dihedral angles were observed. The most striking feature was the extraordinary quality of the electron-density maps clearly locating atomic positions (Fig. 2) even for solvent-exposed side chains. The interdomain connecting peptide occupies a slightly

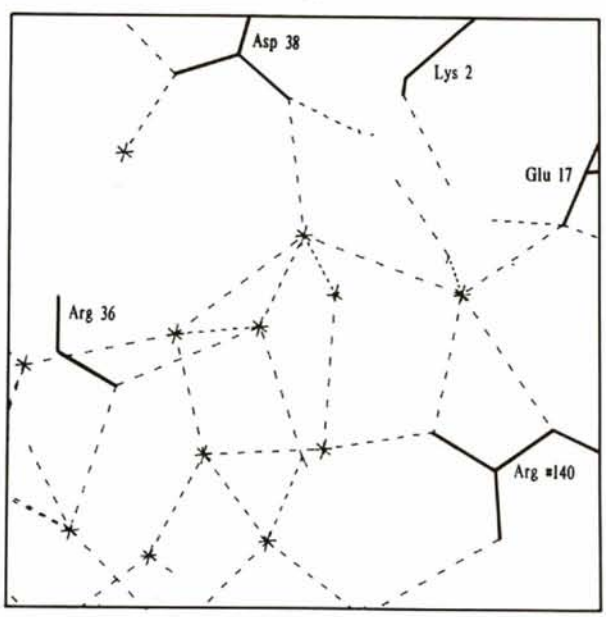
(b)

(c)

Fig. 5. Dot surfaces drawn at van der Waals radii depicting arginine–arginine interactions in γB-crystallin. (*a*) The location of the topologically equivalent pairs of arginine residues between the two domains viewed down the local twofold axis that relates domains. (*b*) The interacting residues Arg58 and Arg168. (*c*) Close interaction of Arg79 with Arg147 showing that the OD1 of Glu#73 salt bridges with the arginines permitting the close approach of the terminal N atoms.



(a)




(b)

Fig. 6. Water organization around charged groups. (*a*) The charged residues which form extensive clusters on the surface of the molecule. Charged groups are shown in yellow. (*b*) The water organization around the free carboxyl of Asp38.

Table 4. *Charged and polar amino-acid side chains*

The contributions of different polar or charged side chains to the hydration of γB-crystallin. The column *Residues* shows the number of each type of residue found in the protein, *Atoms* shows the number of polar or charged atoms in these residues and *Hydrogen bond water* shows the number of hydrogen-bonded water molecules.

| | Residues | Atoms | Hydrogen bond water |
|---|---|---|---|
| Arg | 20 | 60 | 87 |
| Asn | 6 | 12 | 25 |
| Asp | 13 | 26 | 52 |
| Gln | 10 | 20 | 31 |
| Glu | 9 | 18 | 26 |
| Ser | 13 | 13 | 15 |
| Thr | 3 | 3 | 5 |
| Tyr | 14 | 14 | 20 |
| Lys | 2 | 2 | 3 |
| His | 5 | 10 | 14 |
| Totals | 95 | 178 | 278 |

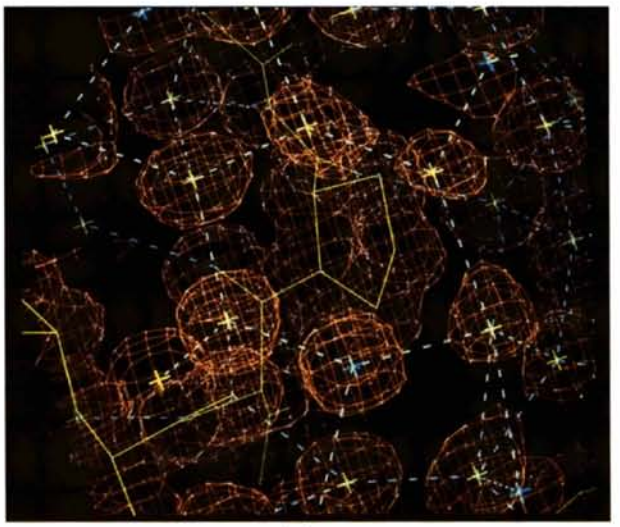
different yet significantly more ordered conformation than that observed in the room-temperature model. For residues 70–90, including the connecting peptide, the r.m.s. differences in coordinates, calculated as above, were 0.34, 0.68 and 0.54 Å for main-chain, side-chain and all protein atoms, respectively. The largest deviations were seen in the side chains of Gln83 and Thr85 (r.m.s. differences 1.65 and 1.33 Å, respectively) and in the main chain at Gly86 (0.67 Å).

### 3.2. *Thermal parameters*

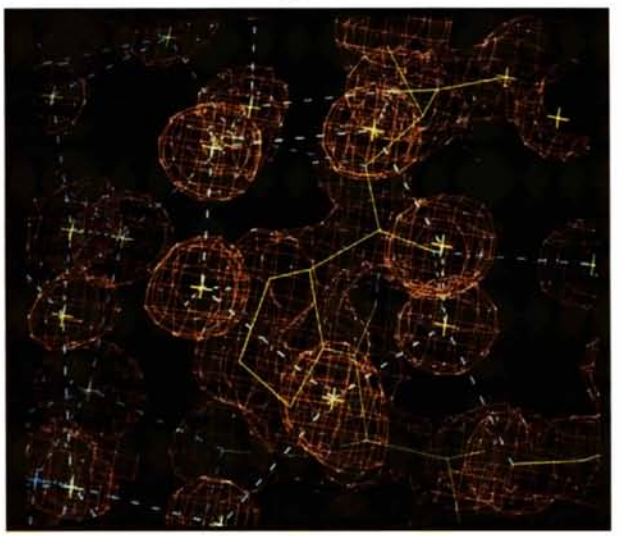
The refined isotropic thermal parameters for γB-crystallin are low. The average isotropic $B$ values for main-chain atoms and side-chain atoms are 5.6 and 11.8 Å$^2$, and the overall average for the 1474 protein atoms is 8.9 Å$^2$. There is variation of $B$ values within the structure, the highest $B$ values are found at the C terminus which appears disordered and in the connecting peptide linking the domains. In general higher thermal parameters are found in surface side chains as is commonly noted. The average values may be compared with those from refinements carried out at room temperature and the 2.0 Å refinement at 150 K. In these cases the average $B$ values for all protein atoms and main-chain protein atoms are 16.1 and 12.5 Å$^2$, and 11.5 and 9.0 Å$^2$, for the room-temperature and low-temperature 2.0 Å models, respectively. It can be seen that the $B$ values have reduced markedly with refinement against very high resolution data. In comparison to the room-temperature data this is probably simply due to the reduction in the amplitudes of thermally enabled atomic vibrations. The large differences from the 2.0 Å model are ascribed to the great improvements in the quality of the electron-density maps using the higher resolution data. As individual atoms are clearly resolved in the electron-density maps, this atomicity defines the limit of atomic electron density and enables the calculation of realistic isotropic $B$ values in the refinement process. This draws a very much clearer picture of conformational disorder superimposed on thermal disorder as was also the case with crambin (Teeter, Roe & Heo, 1993).

### 3.3. *Conformational disorder*

A striking feature of the γB-crystallin–water structure is the low level of conformational disorder. Only four residues Arg9, Ser19, Gln26 and Asp64 were found to have alternative conformations during rebuilding of the model in the later stages of the refinement. In each case the disorder was apparently limited to the alternative



(*a*)



(*b*)

Fig. 7. Water around hydrophobic surfaces. Electron density (at 1$\sigma$ level) and hydrogen-bonded networks around the exposed side chain of Pro82. (*a*) Six-membered ring of hydrogen-bonded water molecules. (*b*) The other side of the proline ring showing a second six-membered ring of hydrogen-bonded water molecules.

occupancy of two discrete sites. Initial occupancies were assigned using relative heights of electron density in the maps and subsequently refined in *RESTRAIN*. For Gln26 (Fig. 3*a*) refined occupancies were close to 60/40% for the alternate conformers and 50/50% for Arg9 (Fig. 3*b*) and Ser19. The second site was only positioned and accepted after checking that they fitted in a sensible manner and made plausible hydrogen bonds to other atoms or solvent molecules. In the case of Gln26 and Arg9 the first site formed solvent-mediated hydrogen-bonded links to a symmetry-related protein molecule. In the second conformer solvent-mediated hydrogen-bonded interactions with the symmetry-related protein molecule were maintained with very similar distances but through an alternative solvent pathway (Fig. 3*b*). Depending upon which site is occupied, the alternate site is occupied by water molecules in good hydrogen-bonding geometry for both possible conformers, introducing a parallel local disorder into the solvent structure.

### 3.4. *Solvent structure*

The use of low-temperature high-resolution data removes many of the obstacles to a detailed analysis of protein–solvent interactions. Exposed side chains are generally much more ordered and the electron density interpreted as water molecules is also better defined. The detailed analysis of solvent structure and interactions formed one of the major aims of this experiment. A crystal density measurement of $1.23 \pm 0.01$ g cm$^{-1}$ (Chirgadze, Nikonov, Garber & Reshetnikova, 1977) and a protein partial specific volume of 0.73 cm g$^{-1}$, indicates $498 \pm 10$ water molecules per asymmetric unit. The high-resolution room-temperature experiment had located a substantial number of water molecules but still less than half of the expected water complement. From an inspection of electron-density maps throughout the course of the refinement we have located and refined at unit occupancy 394 water molecules. No evidence for any sodium or phosphate counterions was found in the solvent structure. The number of water molecules located and refined at unit occupancy represent nearly 80% of the total solvent content per asymmetric unit, but no further significant density, interpretable as water molecules was found. A difference electron-density map was calculated, using structure factors calculated from a model in which one well ordered water molecule was removed from the coordinate list. The deleted water molecule appeared as a peak of $20\sigma$ in the difference map. Any peaks of residual electron density in the (unmodelled) solvent channels appeared with heights less than $3\sigma$. On this basis it may be suggested that the electron-density levels in this region show that water molecules, if located, will be disordered or subject to statistical disorder and present at occupancies lower than 20%. Within the solvent structure 112 of the water molecules with lowest thermal coefficients identified in the 1.47 Å room-temperature refinement are conserved within 1 Å (r.m.s. difference 0.68 Å). The r.m.s. difference in position for all the 230 water molecules found in the low-temperature model was 1.85 Å. The residence times of these preferred hydration sites in the room-temperature crystal are not known (Levitt & Park, 1993) but in the 150 K structure they are essentially static.

Several additional water molecules, but with partial occupancy, are involved in the alternate occupation of the side chains of Arg9 and Gln26. Most of the solvent structure is extremely well ordered, with an average $B$ value of 23.4 Å$^2$. The trend in $B$ values correlates well with the number of protein contacts, decreasing from 27.1 Å$^2$ for one contact to 12.8 Å$^2$ for four protein contacts. Hydrogen-bonding distances depend on the atom types involved in bond formation, average distances being 2.88 and 3.01 Å for main-chain O and N atoms, respectively, and 2.83 and 3.02 Å for the corresponding side-chain atoms.

3.4.1. *Water–protein interactions.* Hydrogen bonds to protein main- and side-chain O and N atoms show distinct average hydrogen-bonding distances ($2.4 < d < 3.5$ Å) of 2.85 and 3.02 Å, respectively, echoing the findings of Roe & Teeter (1993). Protein–water interactions may be grouped into the first hydration shell for those water molecules making at least one hydrogen-bonded interaction with a protein atom and the second or higher hydration shells for those water molecules making hydrogen-bonded interactions only with other water molecules. 268 water molecules are found in the first hydration shell and 126 are distributed in higher hydration shells. The largest proportion of the first hydration shell water molecules make only one protein hydrogen bond ($2.4 < d < 3.5$ Å), but five water molecules make four protein hydrogen bonds and none to other water molecules (Table 3*a*).

Protein to water molecule hydrogen-bonding distances may be classed according to the number of protein contacts and correlated with the number of additional solvent hydrogen bonds, (Table 3*b*). It can be seen that if those water molecules are considered that make four contacts, then the hydrogen-bonding distance decreases as the number of water–water contacts are increased and approaches the value for the second hydration shell (Table 3*b*). The value of 2.91 Å for three protein contacts and one water contact is approaching the average protein–protein hydrogen-bond distance (2.94 Å). In the second and higher hydration shells, the expected trend in $B$ values of the water molecules is seen, the larger the number of hydrogen bonds to further water molecules the lower the average $B$ values (Table 3*c*).

3.4.2. *Solvent around charged residues.* In comparison with other proteins γ-crystallins have numerous ion pairs on the surface (Barlow & Thornton, 1980). Furthermore, an even balance of charges on each of the domains is a conserved feature of γ-crystallin sequences

(Summers *et al.*, 1986). It has been suggested that pairing of charges in γ-crystallins will minimize the interaction with water and so contribute to tight packing in the lens (Blundell *et al.*, 1981; Wistow *et al.*, 1983; Chirgadze, 1992). The structure of a complete water layer over the surface of γB-crystallin in a crystal allows further analysis of the role ion-pair hydration plays in lens structure. The hydration around isolated charged side chains can be compared with those in ion pairs. The bulk of the charged goups in γB-crystallin are involved in ion pairs either intramolecular or between symmetry-related molecules. There are only four (out of 24) free acidic residues in γB-crystallin; Glu17, Asp38, Asp114 and Glu128 (Fig. 4*a*) and there is only one (out of 20) free basic side chain, Arg9. The side chain of Arg9 has a high solvent accessibility and shows conformational disorder in the terminal atoms (Fig. 3*b*) but each alternate conformation has identical water complements.

The Arg9 side chain makes three hydrogen bonds at the NE position (one to protein and two to solvent), two hydrogen bonds to water at the NH1 position and three at the NH2 position. This is an example of a 'fully hydrated' arginine residue. The hydrogen-bonding geometry agrees with hydration patterns found in 16 non-homologous proteins (Thanki, Thornton & Goodfellow, 1988) and with the analysis of hydrogen-bonding clusters (Roe & Teeter, 1993). The hydration of an intramolecular salt bridge (Glu104–Arg89) which is not involved in lattice interactions is shown in Fig. 4(*b*). It can be seen that side chains in ion pairs interact with less water molecules than if they were isolated and fully solvent accessible. Of the 24 acidic side chains in γB-crystallin only the four mentioned above do not form at least one intramolecular ion pair or interaction with either main-chain N atoms or side-chain atoms including ring N atoms of histidines, for example between Glu46 and His53.

The relative contributions of the different hydrophilic amino-acid side chains to the interaction with water in the γB-crystallin crystal lattice is shown in Table 4. The hydration is dominated by the arginine side chains followed by aspartate. In γB-crystallin there are 20 arginine residues. Many are involved in intermolecular interactions either through ion pairs (ten of 20) or intermolecular salt bridges (seven of 20) or hydrogen bonding to solvent. There is, however, considerable flexibility in terms of which groups involved in the interaction occupy the 'head-on' and 'side-on' (Fig. 4*b*) interaction positions for the terminal N atoms (Singh, Thornton, Snarey & Campbell, 1987), although intramolecular interactions are typically mediated through the side-on positions. In contrast with the complementary ion-pair arrangement, side-to-side packing of Arg residues occurs at three sites in the γB-crystallin crystal structure. The Arg58–Arg168 pair is in a topologically equivalent position to the Arg79–Arg147 pair where they are found on either side of the interdomain cleft (Fig. 5*a*). The close juxtaposition of arginine residues inducing a large local electrostatic field must be stabilized either by the local solvent structure (Magalhaes, Maigret, Hoflack, Goities & Scheraga, 1994) or other charged or polar groups. In the first pair, 58–168, the terminal side-chain groups are significantly non-coplanar (Fig. 5*b*) and the closest interactions are van der Waals interactions between the NH2 of 58 and NH2 of 168, at a distance of 3.68 Å. Arg58 has a solvent accessibility of 60 Å$^2$: the NE atom is not accessible and forms no solvent hydrogen bonds, NH2 is partially exposed and forms one solvent hydrogen bond whilst fully exposed NH1 forms two solvent hydrogen bonds. The guanidinium moieties of the topologically equivalent pair, Arg79–Arg147, are nearly coplanar (Fig. 5*c*) with the NH atoms approaching within 3.4 Å, reminiscent of the layers of C atoms in graphite where the delocalized π-orbitals interact through van der Waals forces. Indeed Phe54 has its aromatic plane orientated so that it can be considered as a third layer in conjunction with the two arginine residues. The close approach of charges is stabilized in this case by the formation of a salt bridge involving the NH1 of 147 with Glu73 of a symmetry-related molecule (Fig. 5*c*). Arg47 and Arg152 from two lattice-related molecules may have similar π–π interactions again stabilized by a neighbouring Glu residue (150) forming a bifurcated salt bridge to both arginine residues.

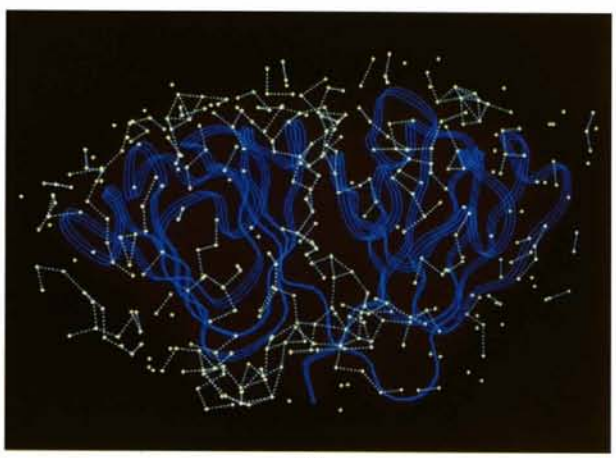
The overall level of exposure of charged groups on the surface of a protein will have considerable effects on the organization of solvent molecules around the protein. More exposed charges will tend to enforce local ordering of the immediate solvent environment. The high degree of ion-pairing interactions, specifically intramolecular ion pairing, will reduce the total number of interaction sites for water molecules. Extensive pairing of charges is seen in the γB-crystallin molecule in the clustering of charged groups in distinct regions of the protein surface. This clustering of acidic and basic groups will reduce the overall level of charge on the surface of the protein. The clustering forms extensive bands over the surface of the molecule (Fig. 6*a*) and extends to include the locally planar charged residues previously reported by Chirgadze (1992). Within these clusters there are extensive networks of intramolecular charge pairing either directly or through water bridging (Fig. 6*b*). In most cases the water bridging between ions involves only one or two water molecules. Part of the charge cluster, centred upon Asp38 in a highly solvent-accessible region of the molecule, forms extensive hydrogen-bonded water bridges between charged groups forming a complex local network of water molecules centred upon four interconnected pentagonal rings of water molecules (Fig. 6*c*).

These molecular observations give an indication of how γ-crystallins might interact in the different stages of lens maturation. In hydrated young lenses γ-crystallins

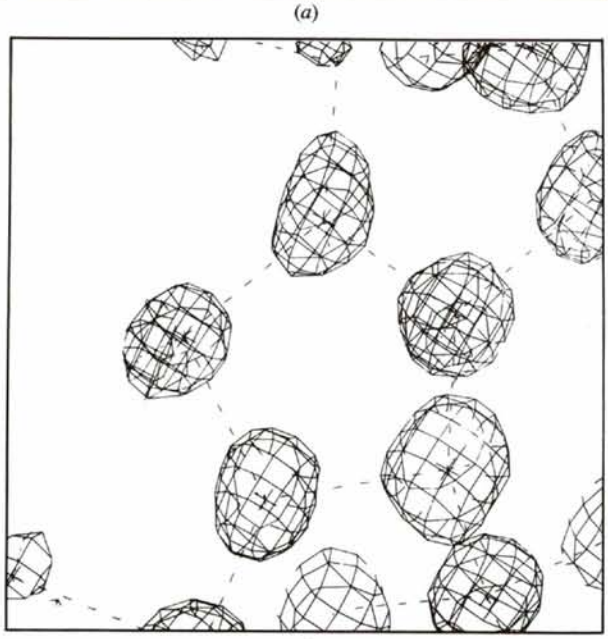
will be predominantly surrounded by water with the numerous intramolecular ion pairs making a major contribution to the protein–water interaction similar to that shown in Fig. 4(*b*) (Glu104–Arg89). A delicate balance of protein–water *versus* protein–protein interaction is in play in the liquid-like lens as lowering the temperature causes light scattering because of separation into phases of uneven protein concentration (Benedek *et al.*, 1979; Blundell *et al.*, 1983). In the centre of older lenses, the protein is glass-like, and the protein concentration approaches that of crystals. Here the ion pairs will reduce the water interaction further by making intermolecular protein contacts.

3.4.3. *Hydrophobic hydration.* A significant fraction of the protein surface not involved in lattice contacts is hydrophobic. γB-crystallin's exposed surface area is 44% hydrophobic/56% hydrophilic in the crystal and this is less hydrophobic than other protein crystals such as crambin, myoglobin and carboxypeptidase which have 60/40%, 57/43% and 55/45% hydrophobic and hydrophilic accessible surface area, respectively (Teeter, Roe & Heo, 1993). The average interaction distance between all hydrophobic atoms and first-shell water molecules ($2.8 < d < 4.5$ Å) is 4.19 Å, and the average $B$ factor of water molecules involved in these interactions is 18 Å$^2$. This is in agreement with a distance of 4.0 Å calculated from a survey of solvent hydrophobic interactions in 24 protein structures (Goodfellow, Thanki & Thornton, 1993; Walshaw & Goodfellow, 1993).

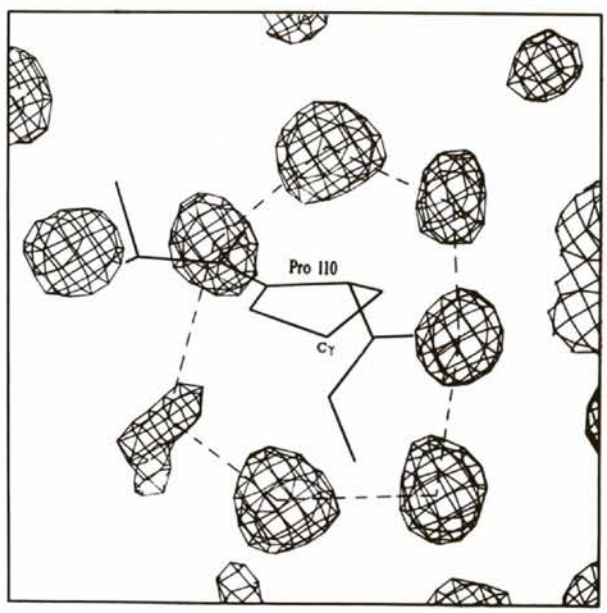
In an analogous manner to the hydration of hydrophilic atoms the organization of solvent molecules around hydrophobic atoms will depend critically upon their environment and residue types. An analysis of solvent interactions with phenylalanine residues shows an



(*a*)



(*b*)




(*c*)

Fig. 8. Organization of water molecules in the γB-crystallin protein–water complex. (*a*) Networks of hydrogen-bonded water molecules. The largest network, including water molecules through the interdomain cleft, comprises 64 water molecules. (*b*) Electron density contoured at the 1$\sigma$ level in the solvent region describing a fivefold ring of water molecules. (*c*) A large water ring around exposed hydrophobic groups.

average interaction distance of 3.94 Å for 114 contacts with water molecules with an average $B$ value 19 Å$^2$. For proline residues the corresponding distance is 3.95 Å for 122 interactions with water molecules of average $B$ value 20 Å$^2$.

Specific examples yield an insight into the organization of solvent around hydrophobic groups. Pro82 is exposed, being on the edge of the wide crevice between the two domains of the molecule. The local water structure around Pro82 forms an extensive three-dimensional cage with 33 contacts of average distance 3.98 Å (average $B$ value 18 Å$^2$). Within this cage the closest contacts to hydrophobic atoms are 3.6 Å, close to the van der Waals contact distance. These closest contacts are arranged as a halo of water molecules following the path of the ring side chain (Figs. 7*a* and 7*b*). Phe71 has a high solvent-accessible area, 130 Å$^2$. In this case the solvent is again organized as a complex cage around the side chain, for 29 contacts within the range specified above the average contact distance is 3.87 Å (average $B$ value 15 Å$^2$), with nearest contacts at or very close to the van der Waals distance. Three of the closest water molecules lie close to the plane of the phenylalanine ring and approximately perpendicular to the middle of the intercarbon bonds, close to that suggested as a common theme for organization around ring side chains (Walshaw & Goodfellow, 1993).

### 3.5. *Solvent organization*

With almost 80% of the complete asymmetric unit complement of water refined at unit occupancy and with low disorder the model presents an excellent opportunity to gain insight into the overall solvent organization around a protein molecule in a situation of great biological relevance. The solvent forms extensive hydrogen-bonded networks, the largest including 64 water molecules in a single network, and two networks of over 40 water molecules are also seen. Water molecules are involved in intermolecular contacts, forming hydrogen-bonded bridges between protein molecules and in intramolecular interactions, the most prominent of these being the small cluster of water molecules in the interdomain cleft, forming part of the largest water network (Fig. 8*a*). It has been suggested that the water-filled cavities between protein domains play a role in molecular motions (Frey, 1993).

Solvent hydrogen bonding depends on the hydration shell the water molecule originates in and is modulated by the number and type of protein hydrogen bonds. In the second hydration shell the average O···O hydrogen-bonding distance is that observed in 'bulk' water. The hydrogen-bonding distances of water to protein are clearly modulated by the relative number of water–water hydrogen bonds and water–protein hydrogen bonds. On average a water molecule with three protein hydrogen bonds and one water hydrogen bond has a longer hydrogen-bonding distance than a water molecule with three water hydrogen bonds and one protein hydrogen bond (Table 3*b*) even though the $B$ value of the latter will, in general, be larger.

Table 5. *Solvent networks*

The columns headed *Protein + Water* and *Water* refer to the total number of rings involving protein and water molecules and water molecules only, respectively. One seven-membered ring was observed which comprised only of water molecules (Fig. 8*c*).

| | Protein + Water | Water |
|---|---|---|
| Three-membered | 99 | 37 |
| Four-membered | 81 | 23 |
| Five-membered | 93 | 32 |
| Six-membered | 4 | 4 |

The solvent forms extensive hydrogen-bonded networks accounting for 100% of the solvent-accessible surface of the protein. The largest network contains 64 water molecules with others over 40. The water networks are based on a number of local, hydrogen-bonded ring geometries.

In the early stages of the refinement it was clear that the solvent structure in the crystal was extremely well ordered and patterns of organization were sought. In common with previous room-temperature studies a unit of organization seemed to be a pentagonal assembly of hydrogen-bonded water molecules (Neidle, Berman & Shieh, 1980; Teeter, 1984; Karplus & Schultz, 1987; Jeffrey & Saenger, 1991). This local motif of organization was widely seen in the 150 K structure of γB-crystallin involving both clusters of water molecules (Fig. 8*b*) and water molecules hydrogen bonded to protein atoms. Indeed as the alternate conformers were identified at Gln26 two such five-membered rings were formed depending upon which subsite was occupied (Fig. 3*a*). As the refinement progressed, however, other patterns of organization of solvent molecules around the protein molecule became apparent. Different water-ring structures were observed depending on the local environment. Assemblies of three, four, five and six water molecules in ring structures are commonly observed. All ring assemblies are apparently centred upon a common tetrahedral arrangement of hydrogen bonds about each water molecule. The organization of water molecules into local rings of varying geometry extends to the inclusion of exposed polar atoms which also contribute atoms to local surface networks (Table 5).

In the environment around hydrophobic regions of the protein surface extensive three-dimensional water cages are formed, screening the hydrophobic patches. This is most evident in the vicinity of the exposed hydrophobic residue Pro82. Here the solvent is seen to be organized in a number of ring structures (Fig. 7). However, on either side of this exposed proline residue six-membered rings are evident. This may constitute a local method of forming the most extensive cage structure, enclosing a large hydrophobic group, using the least number of water molecules. Six-membered rings are also seen in solvent

structure near the exposed residue, Phe71 (Fig. 8c). In both these cases the six-membered rings are formed in solvent regions where no polar or charged protein atoms contribute to the hydrogen-bonding pattern. In the absence of such polar groups the local water structure mimics more closely that expected of pure water such as in the structure of ice.*

We are grateful for financial support from the Medical Research Council, CCLRC Daresbury Laboratory and Keele University.

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1AMM, R1AMMSF). Free copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Reference: HE0126).

## References


Barlow, D. J. & Thornton, J. M. (1983). *J. Mol. Biol.* **168**, 867–885.

Bax, B., Lapatto, R., Nalini, V., Driessen, H., Lindley, P. F., Mahadevan, D., Blundell, T. L. & Slingsby C. (1990). *Nature (London)*, **347**, 776–780.

Benedek, G. B., Clark, I. J., Serrallach, E. N., Young, C. Y., Mengel, L., Sauke, T., Bagg, A. & Benedek, K. (1979). *Philos. Trans. R. Soc. London Ser. A*, **293**, 329–340.

Bloemendal, H. & de Jong, W.W. (1991). *Prog. Nucleic Acid Res. Mol. Biol.* **41**, 259–281.

Blundell, T., Lindley, P., Miller, L., Moss, D., Slingsby, C., Tickle, I. Turnell, B., Wistow, G. (1981). *Nature (London)*, **289**, 771–777.

Blundell, T. L., Lindley, P. F., Miller, L. R., Moss, D. S., Slingsby, C., Turnell, W. G. & Wistow, G. (1983). *Lens Res.* **1**, 109–131.

Broide, M. L., Berland, C. R., Pande, J., Ogun, 0. 0. & Benedek, G. B. (1991). *Proc. Natl Acad. Sci. USA*, **88**, 5660–5664

Carlisle, C. H., Lindley, P. F., Moss, D. S. & Slingsby, C. (1977). *J. Mol. Biol.* **110**, 417–419.

Chirgadze, Y. N. (1992). *Mol. Biol.* **26**, 940–944.

Chirgadze, Y. N., Nikonov, S. V., Garber, M. B. & Reshetnikova, L. S. (1977). *J. Mol. Biol.* **110**, 619–624.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Driessen, H. P. C., Haneef, I., Harris, G. W., Howlin, B., Khan, G. & Moss, D. S. (1989). *J. Appl. Cryst.* **22**, 510–516.

Fox, G. C. & Holmes, K. C. (1966). *Acta Cryst.* **20**, 886–891.

Frey, M. (1993). *Water and Biological Macromolecules*, edited by E. Westhof, pp. 98–147. London: Macmillan Press.

Goodfellow, J. M., Thanki, N. & Thornton, J. M. (1993). *Water and Biological Macromolecules*, edited by E. Westhof, pp. 63–97. London: Macmillan Press.

Harding, J. (1991) *Cataract: Biochemistry, Epidemiology and Pharmacology*. London: Chapman & Hall.

Hartmann, H., Parak, F., Steigemann, W., Petsko, G., Ponzi, D. & Frauenfelder, H. (1982), *Proc. Natl Acad. Sci. USA*, **79**, 4967–4971.

Hope, H. (1988). *Acta Cryst.* B**44**, 22–26.

Horwitz, J. (1992). *Proc. Natl Acad. Sci. USA*, **89**, 10449–10453.

Jeffrey, G. A. & Saenger, W. (1991). *H-bonding in Biological Structures*. Berlin: Springer-Verlag.

Jones, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.

Karplus, P. A. & Faerman, C. (1994). *Curr. Opin. Struct. Biol.* **4**, 770–776.

Karplus, P. A. & Schulz, G. E. (1987). *J. Mol. Biol.* **195**, 701–729.

Klemenz, R., Fröhli, E., Steiger, R. H., Schafer, R. & Aoyama, A. (1991). *Proc. Natl Acad. Sci. USA*, **88**, 3652–3656.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Levitt M. & Park, B. (1993). *Structure*, **1**, 223–226.

Lindley, P., Najmudin, S., Bateman, O., Slingsby, C., Myles, D., Kumaraswamy, V.S. & Glover, I. (1993). *J. Chem. Soc. Faraday Trans.* **89**, 2677–2682.

Lubsen, N. H., Aarts, H. J. M., & Schoenmakers, J. G. G. (1988). *Prog. Biophys. Mol. Biol.* **51**, 47–76.

Magalhaes, A., Maigret, B., Hoflack, J., Goities, J. N. F. & Scheraga, H. A. (1994). *J. Protein Chem.* **13**, 195–215.

Najmudin, S., Nalini, V., Driessen, H. P. C., Slingsby, C., Blundell, T. L., Moss, D. S. Lindley, P. F. (1993). *Acta Cryst.* D**49**, 223–233.

Neidle, S., Berman, H. M. & Shieh, H. S. (1980). *Nature (London)*, **288**, 129–133.

Nyburg, J. & Wonacott, A. J. (1977). In *The Rotation Method in Crystallography*, edited by U. Arndt & A. J. Wonacott. Amsterdam: North Holland,

Philipson, B. (1969). *Invest. Ophthalmol.* **8**, 258–270.

Roe, S. M. & Teeter, M. M. (1993). *J. Mol. Biol.* **229**, 419–427.

Simpson, A., Bateman, O., Driessen, H., Lindley, P., Moss, D., Mylvaganam, S., Narebor, E. & Slingsby, C. (1994). *Nature Struct. Biol.* **1**, 724–734.

Singh, J., Thornton, J. M., Snarey, M. & Campbell, S. F. (1987). *FEBS Lett.* **224**, 161–171.

Slingsby, C. (1985). *Trends Biol. Sci.* **10**, 281–284.

Summers, L. J., Slingsby, C., Blundell, T. L., Den Dunnen, J. T., Moormann, R. J. M. & Schoenmakers, J. G. G. (1986). *Exp. Eye Res.* **43**, 77–92.

Tardieu, A., Vérétout, F., Krop, B. & Slingsby, C. (1992). *Eur. Biophys. J.* **21**, 1–12.

Teeter, M. M. (1984). *Proc. Natl Acad. Sci. USA*, **81**, 6014–6018.

Teeter, M. M., Roe, S. M. & Heo, N. H. (1993). *J. Mol. Biol.* **230**, 292–311.

Thanki, N., Thornton, J. M. & Goodfellow, J. M. (1988). *J. Mol. Biol.* **202**, 637–657.

Walshaw, J. & Goodfellow, J. M. (1993). *J. Mol. Biol.* **231**, 392–414.

Wannermacher, C. F. & Spector, A. (1968). *Exp. Eye Res.* **7**, 623–635.

Wistow, G., Turnell, B., Summers, L., Slingsby, C., Moss, D., Miller, L., Lindley, P. & Blundell, T. (1983). *J. Mol. Biol.* **170**, 175–202.

Wistow, G. J. & Piatigorsky, J. (1988). *Annu. Rev. Biochem.* **57**, 479–504.